AI and the Art of Subtle Control

Muskula Rahul

I didn't grow up fearing computers. For my generation, technology wasn't something terrifying—it was our constant companion. Yet those early "computerphobia" concerns weren't entirely misguided. While people feared replacement, what actually happened was something more profound: machines didn't replace us—they **integrated seamlessly into the fabric of our daily existence**.

Today, AI presents a challenge more nuanced than any science fiction narrative. Not killer robots. Not superintelligent overlords.

AI is already reshaping our cognition, curating our reality, and guiding our beliefs—silently, systematically, and often beyond our awareness.

The Architecture of Influence: Behavioral Engineering at Scale

Forget doomsday scenarios. The immediate danger isn't artificial *super*intelligence—it's **artificial influence**. Today's algorithms don't require consciousness—just pervasiveness and persuasiveness.

Major platforms—Facebook, Instagram, TikTok, YouTube—track your clicks, hesitations, emotional triggers, and attention patterns with millisecond precision. They deliver content optimized for engagement—triggering emotions, reinforcing identity markers, and steering opinions.

This is **behavioral engineering at scale**.

Twitter's internal research revealed their algorithm amplified right-leaning political content in six out of seven countries, not by design but because divisive content generates higher engagement [9]. Similarly, Facebook's researchers discovered their recommendation systems naturally gravitate toward content that triggers outrage—the emotion driving the highest engagement metrics.

On YouTube, Stanford researchers discovered algorithmic "radicalization pathways" that directed users toward increasingly extreme content [1]. The platform's recommendation engine, responsible for over 70% of viewing time, creates what researchers call "filter bubbles" that reinforce existing beliefs and gradually shift viewpoints toward more extreme positions.

These algorithms don't just know *what* content you consume—they measure how long your eyes linger on specific elements, which emotional triggers prompt you to share, and what psychological levers keep you engaged. This knowledge is then weaponized to extend your session time and maximize advertising revenue.

The Invisible Experiment You Never Consented To

These systems construct psychological profiles more detailed than most therapists. Without explicit consent, they map your political leanings, insecurities, and behavioral patterns.

You aren't selecting what enters your awareness. The algorithm is.

And its priorities aren't truth or public good—they're **attention, engagement, and profit**.

The Facebook Files, revealed by whistleblower Frances Haugen, showed Meta ignored warnings that its algorithm promoted divisiveness. Instagram's negative effects on teens were known internally, but growth took precedence [6]. Internal documents showed executives were aware their algorithm exacerbated body image issues among teenage girls, yet consistently prioritized engagement metrics over mental health concerns.

This manipulation extends beyond social media. Dating apps employ variable reward mechanisms—similar to slot machines—to keep users swiping instead of matching. Streaming services conduct thousands of A/B tests to determine which thumbnails and auto-playing previews will maximize your viewing time.

The psychological techniques deployed weren't developed accidentally. They're the result of billions in research, applying decades of behavioral psychology to digital environments, optimized through continuous experimentation on billions of unwitting users.

From Personalized Bubbles to Societal Fragmentation

Cambridge Analytica harvested data from 87 million Facebook users to craft targeted psychological operations during Brexit and the 2016 U.S. election [5]. This wasn't simple advertising—it was precision-targeted emotional manipulation using psychographic profiles derived from users' digital footprints.

In Myanmar, algorithmic amplification of hate speech on Facebook played a significant role in inciting violence against the Rohingya minority [4]. UN investigators characterized the platform as having played a "determining role" in the crisis that led to over 700,000 refugees fleeing violence.

MIT studies found misinformation spreads **6x faster** than truthful content—due to algorithmic bias toward emotionally charged content [11]. This speed differential creates a fundamental asymmetry: falsehoods outpace corrections, emotions overpower facts, and outrage drowns out nuance.

The polarization isn't merely ideological—it's epistemic. Different segments of society now operate with fundamentally different sets of "facts," making democratic consensus increasingly difficult. A 2023 Pew Research study found Americans exposed to different media ecosystems hold not just different opinions, but entirely different perceptions of basic reality.

This fragmentation isn't accidental—it's an emergent property of engagement-optimized systems that profit from capturing attention through emotional arousal. When platforms optimize for engagement rather than understanding, they naturally drive toward division rather than consensus.

China's Digital Authoritarianism: The Blueprint for Control

In China, surveillance is pervasive and growing more sophisticated:

- **Over 626 million AI-enabled cameras** (one for every two citizens)
- Facial recognition that identifies people in crowds of 50,000+
- AI tracks online behavior, purchases, and relationships
- Social credit systems reward or penalize citizens based on conformity

This system doesn't just observe—it **modifies behavior**, rewarding compliance and punishing dissent without physical coercion. It represents the most advanced implementation of algorithmic governance in history.

The Chinese model demonstrates how AI can enable unprecedented levels of social control without traditional authoritarian methods. Critics cannot be silenced if algorithms ensure they're never heard. Dissent doesn't need to be crushed if it can be preemptively discouraged through social credit penalties.

What makes this approach particularly concerning is its exportability. China actively promotes its surveillance technologies to other governments. At least 80 countries have now imported Chinese surveillance systems, creating what experts call ”digital authoritarianism as a service.”

The danger isn't that machines will become conscious and rebel—it's that they'll work exactly as designed, optimizing for control and conformity at the expense of human autonomy.

Next-Generation Tools: Manipulation at Scale

AI advances are enhancing manipulation capabilities at an accelerating pace:

- Voice cloning tools like ElevenLabs can recreate speech from seconds of audio with near-perfect accuracy
- Text-to-video generators (Runway, Sora) can fabricate photorealistic fake events indistinguishable from reality
- Language models generate believable, biased content at scale, creating thousands of personalized messages
- Emotion recognition algorithms claim to read psychological states from facial expressions and voice patterns
- Cross-modal AI can generate coordinated disinformation across text, audio, images, and video simultaneously

In 2024, deepfake robocalls imitating President Biden attempted to suppress votes in New Hampshire [10]. The calls sounded so authentic that many voters were confused about voting procedures, demonstrating how these technologies directly undermine democratic processes.

India's 2024 elections saw AI-generated political deepfakes targeting religious groups before verification could catch up [8]. One fabricated video showing a politician insulting a religious minority spread to millions before being identified as synthetic, triggering localized violence in several communities.

These technologies create what researchers call an "authenticity crisis" — when seeing and hearing is no longer believing, society loses crucial epistemological anchors. The window of time between a new synthetic media capability and effective detection creates periods of extreme vulnerability.

What makes this particularly concerning is the asymmetry: creating convincing synthetic media requires far less resource and expertise than detecting or mitigating its effects. A single individual can now produce disinformation at industrial scale.

The Decision Infrastructure: Algorithms as Gatekeepers

AI increasingly determines who gets hired, loans, education, or parole — functioning as an invisible layer of decision-making infrastructure that shapes opportunities for billions.

A 2023 report found **over 80% of Fortune 500 companies** used AI in hiring, yet fewer than 15% audited these systems for bias [7]. Amazon scrapped its AI recruiting tool after discovering it penalized résumés with the word "women's" [2] — having learned from historical hiring patterns that reflected gender bias in the tech industry.

In lending, algorithms determine credit worthiness based on thousands of data points beyond traditional credit scores. These systems often reproduce historical patterns of discrimination while adding an unassailable veneer of objectivity through mathematical complexity.

Criminal justice has seen similar concerns. Predictive policing algorithms determine which neighborhoods receive additional police presence, often reinforcing existing patterns of over-policing in minority communities. Recidivism prediction tools like COMPAS have been shown to produce racially disparate outcomes despite similar criminal histories.

What makes algorithmic gatekeeping particularly troubling is the combination of opacity, scale, and lack of accountability. When an individual loan officer shows bias, the impact is limited. When an algorithm deployed across an entire industry shows bias, it affects millions. And unlike human decision-makers who can be questioned, algorithmic systems often function as unaccountable black boxes.

The Attention Economy: Cognitive Capitalism's End Game

Perhaps the most fundamental shift AI has enabled is the industrialization of attention capture and manipulation. Our cognitive bandwidth — what we notice, care about, and remember — has become the most valuable commodity in the digital economy.

This "attention economy" represents a profound shift in how capitalism functions. Traditional capitalism commodified physical labor; cognitive capitalism commodifies awareness itself. AI-powered systems have made this extraction exponentially more efficient.

Research indicates the average American spends over 7 hours daily on digital media, with attention increasingly fragmenting across multiple devices. During this time, machine learning systems continuously optimize content to maximize engagement, creating what researchers call "attentional labor" — unpaid cognitive work that generates value for platforms.

The consequences for human psychology are profound:

- Studies show sustained decreases in attention spans
- Increases in anxiety and depression

- Fundamental changes in how people process information

A Harvard study found people now commonly read in an ”F-pattern” — scanning headlines and first sentences rather than processing content deeply.

These changes represent not just individual psychological shifts but a societal-level transformation in how we relate to information, each other, and reality itself. When attention becomes the primary battlefield of economic competition, the result is an increasingly frantic, fragmented, and manipulated information environment.

Beyond Technological Solutionism: Social and Political Dimensions

Technology alone cannot solve problems that are fundamentally social and political. While better algorithms matter, equally important are the power structures, economic incentives, and cultural values that shape how technologies are developed and deployed.

Meaningful reform requires addressing four interrelated dimensions:

1. **Technical systems** — how algorithms and platforms function
2. **Economic models** — how digital businesses create and extract value
3. **Governance frameworks** — how we collectively regulate technological development
4. **Cultural narratives** — how we understand and relate to technology

Progress requires coordinated movement across all four dimensions. Better technical systems will fail if economic incentives reward exploitation. Strong regulations will falter if cultural narratives prioritize convenience over autonomy.

This means engaging with AI as a political issue, not merely a technical one. Questions of who develops AI, who benefits from it, and who bears its costs are fundamentally questions about power and justice — questions that cannot be answered through engineering alone.

A New Digital Social Contract

What we need is a new digital social contract — an updated understanding of the rights, responsibilities, and relationships that should govern our increasingly AI-mediated society.

This contract should include:

- **Algorithmic accountability**: Systems that make important decisions must be explainable, contestable, and remediable
- **Attention sovereignty**: People have the right to control what claims their attention
- **Information integrity**: Society has a collective interest in maintaining a factual information commons
- **Distributive justice**: The benefits of automation and AI must be widely shared
- **Cognitive liberty**: People have the right to mental self-determination free from manipulation

Such a contract requires reimagining not just individual technologies but the broader social, economic, and political systems in which they're embedded. It means treating digital infrastructure as essential public goods rather than private extraction opportunities.

As AI becomes increasingly integrated into every aspect of society, the decisions we make now will shape not just individual products or companies, but the future of human autonomy itself.

Reimagining AI: Tools for Empowerment, Not Exploitation

AI isn't inherently dangerous. It can drive medical breakthroughs, solve climate problems, and aid accessibility. Recent advances in protein folding prediction, climate modeling, and assistive technologies demonstrate its tremendous positive potential.

But we must rethink how we build and deploy it:

- **Transparency** over opacity
- **Human agency** over algorithmic nudging
- **Wellbeing** over profit
- **Distributed benefit** over corporate monopoly

Let AI be a **cognitive partner**, not a manipulative master. This requires moving beyond purely technical considerations to address the economic incentives, power dynamics, and social contexts in which AI systems operate.

Technical solutions alone won't suffice. We need comprehensive reform of the digital economy:

- Data rights legislation giving individuals genuine control over their information
- Mandatory algorithmic impact assessments before deployment of high-risk systems
- Public interest alternatives to private surveillance platforms
- Digital literacy education equipping citizens to navigate algorithmically-mediated spaces
- Antitrust enforcement to break up platform monopolies

Most fundamentally, we need to shift from an attention economy to what some scholars call a ”respect economy” — one that treats human attention as sacred rather than extractable.

Pioneers of Human-Centered AI

Several initiatives showcase ethical alternatives:

- **Center for Humane Technology** works to realign technology with humanity's best interests
- **Signal** provides privacy-first communication without data harvesting
- **DuckDuckGo** offers search without tracking or filter bubbles
- **Mozilla's Common Voice** creates open speech datasets that democratize voice AI

- **Partnership on AI** brings together diverse stakeholders for ethical AI development
- **Algorithm Justice League** works to highlight and mitigate algorithmic bias
- **The Alignment Research Center** focuses on ensuring AI systems remain beneficial as they grow more powerful

Yoshua Bengio's AI Commons and Finland's nationwide AI literacy campaign show how we can democratize AI [3]. These initiatives demonstrate that alternative paths exist — ones that harness AI's potential while respecting human autonomy and wellbeing.

Community-owned platforms like Mastodon demonstrate how social technologies can function without surveillance-based business models. Cooperatively developed AI systems show how beneficial capabilities can be developed outside of commercial incentives.

These pioneers remind us that there's nothing inevitable about AI's current trajectory. The technologies we build reflect our values and priorities. Different design choices lead to different futures.

Conclusion

The question is not ”what can AI do?” but **”what values does it serve?”**

> Will AI empower the many—or manipulate the many for the gain of a few?

We need AI that:

- Illuminates rather than manipulates
- Enhances agency instead of replacing it
- Serves the public, not just platforms
- Distributes power rather than concentrating it
- Respects cognitive liberty and autonomy

The existential threat isn't that machines will overthrow humanity—it's that they'll perfect the exploitation of human vulnerability. The battle isn't against conscious machines but against dehumanizing systems.

This isn't a call for technophobia. It's an invitation to imagine and build AI that genuinely serves human flourishing—AI that amplifies our best qualities rather than exploiting our weaknesses.

The path forward requires not just technical innovation but moral imagination. We need to ask not just what's possible, but what's desirable—not just what AI can do, but what it should do.

The real frontier isn't artificial intelligence, but artificial wisdom—systems that embody not just computational capability but human values, ethical judgment, and respect for the irreducible complexity of human experience.

The future of AI is still unwritten. Let's ensure it's a story of liberation rather than control.

References

[1] Ribeiro et al. “Auditing Radicalization Pathways on YouTube”. In: (2020). URL: https://dl.acm.org/doi/10.1145/3351095.3372879.

[2] “Amazon scraps secret AI recruiting tool that showed bias against women”. In: (2018). URL: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

[3] Yoshua Bengio. “AI Commons: Democratizing AI for the benefit of all”. In: (2023). URL: https://aicommons.org.

[4] “Facebook admits it was used to incite violence in Myanmar”. In: (2018). URL: https://www.theguardian.com/technology/2018/nov/06/facebook-admits-it-has-not-done-enough-to-quell-hate-in-myanmar.

[5] *Facebook–Cambridge Analytica Data Scandal*. 2018. URL: https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal.

[6] Frances Haugen. “The Facebook Files”. In: (2021). URL: https://www.wsj.com/articles/the-facebook-files-11631713039.

[7] “How AI is shaping the hiring process”. In: (2023). URL: https://hbr.org/2023/04/how-ai-is-shaping-the-hiring-process.

[8] “India’s election and the deepfake disinformation dilemma”. In: (2024). URL: https://www.lemonde.fr/en/pixels/article/2024/05/21/india-s-general-election-is-being-impacted-by-deepfakes_6672168_13.html.

[9] Twitter Responsible ML. “Algorithmic amplification of politics on Twitter”. In: *Twitter Blog* (2021). URL: https://arxiv.org/abs/2110.11010.

[10] “Political consultant charged for AI-generated robocalls mimicking Biden”. In: (2024). URL: https://apnews.com/article/biden-robocalls-ai-new-hampshire-charges-fines-9e9cc63a71eb9c78b9bb0d1ec2aa6e9c

[11] Aral Vosoughi Roy. “The spread of true and false news online”. In: *Science* (2018). URL: https://www.science.org/doi/10.1126/science.aap9559.